**Statistics**

<https://en.wikipedia.org/wiki/Statistics>

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.[1] In applying statistics to, e.g., a scientific, industrial, or societal problem, it is conventional to begin with a statistical population or a statistical model process to be studied. Populations can be diverse topics such as "all persons living in a country" or "every atom composing a crystal". Statistics deals with all aspects of data including the planning of data collection in terms of the design of surveys and experiments.[1]

When census data cannot be collected, statisticians collect data by developing specific experiment designs and survey samples. Representative sampling assures that inferences and conclusions can safely extend from the sample to the population as a whole. An experimental study involves taking measurements of the system under study, manipulating the system, and then taking additional measurements using the same procedure to determine if the manipulation has modified the values of the measurements. In contrast, an observational study does not involve experimental manipulation.

Two main statistical methodologies are used in data analysis: descriptive statistics, which summarizes data from a sample using indexes such as the mean or standard deviation, and inferential statistics, which draws conclusions from data that are subject to random variation (e.g., observational errors, sampling variation).[2] Descriptive statistics are most often concerned with two sets of properties of a distribution (sample or population): central tendency (or location) seeks to characterize the distribution's central or typical value, while dispersion (or variability) characterizes the extent to which members of the distribution depart from its center and each other. Inferences on mathematical statistics are made under the framework of probability theory, which deals with the analysis of random phenomena.

A standard statistical procedure involves the test of the relationship between two statistical data sets, or a data set and a synthetic data drawn from idealized model. An hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis of no relationship between two data sets. Rejecting or disproving the null hypothesis is done using statistical tests that quantify the sense in which the null can be proven false, given the data that are used in the test. Working from a null hypothesis, two basic forms of error are recognized: Type I errors (null hypothesis is falsely rejected giving a "false positive") and Type II errors (null hypothesis fails to be rejected and an actual difference between populations is missed giving a "false negative"). Multiple problems have come to be associated with this framework: ranging from obtaining a sufficient sample size to specifying an adequate null hypothesis.[citation needed]

Measurement processes that generate statistical data are also subject to error. Many of these errors are classified as random (noise) or systematic (bias), but other important types of errors (e.g., blunder, such as when an analyst reports incorrect units) can also be important. The presence of missing data and/or censoring may result in biased estimates and specific techniques have been developed to address these problems.

Statistics can be said to have begun in ancient civilization, going back at least to the 5th century BC, but it was not until the 18th century that it started to draw more heavily from calculus and probability theory. Statistics continues to be an area of active research, for example on the problem of how to analyze Big data.